Learning with Synaptic Nonlinearities in a Coupled Oscillator Model of Olfactory Cortex*

Bill Baird

Depts. Mathematics and Molecular and Cell Biology,
U.C.Berkeley, Berkeley, Ca. 94720

Abstract

A simple network model of olfactory cortex, which assumes only minimal coupling justified by known anatomy, can be analytically proven to function as an associative memory for oscillatory patterns. The network has explicit excitatory neurons with local inhibitory interneuron feedback that forms a set of nonlinear oscillators coupled only by long range excitatory connections. Using a local Hebb-like learning rule for primary and higher order synapses at the ends of the long range connections, the system can learn to store the kinds of oscillation amplitude and phase patterns observed in olfactory and visual cortex. The network can be truely self-organizing because a synapse can modify itself according to it's own pre and postsynaptic activity during stimulation by an input pattern to be learned. The neurons of the neuron pools modeled here can be viewed as operating in the linear region of the usual sigmoidal *axonal* nonlinearity, and multiple memories are stored instead by the learned *synaptic* nonlinearities.

1 Introduction

We report recent results of work (discussed fully in [1, 3, 2, 4]) which seeks to narrow the gap that exists between physiologically detailed network models of real vertebrate cortical memory systems and analytically understood "artificial" neural networks for associative memory. The secondary olfactory sensory cortex known as prepyriform cortex is thought to be one of the clearest cases of a real biological network with associative memory function [9]. The fact that it exhibits oscillatory dynamics makes application of the usual connectionist notions difficult. A network that is a caricature of the anatomy of the prepyriform cortex has been analyzed, and Hebbian learning rules derived which can analytically guarantee the storage of $N/2$ oscillatory memory patterns as the only attractors of an N node network[1, 2]. We have shown that this model is also capable of storing complicated multifrequency spatio-temporal trajectories[3], and argue that it may serve as a model of memory for sequences of actions in premotor cortex[4].

*To appear in *Analysis and Modeling of Neural Systems*, Frank H. Eeckman, editor. Supported by AFOSR-87-0317. It is a pleasure to acknowledge the support of Walter Freeman and invaluable assistance of Morris Hirsch.

Patterns of 40 to 80 Hz oscillation have been observed in the large scale activity (local field potentials) of olfactory cortex [6, 7] and visual neocortex [8], and shown to predict the olfactory and visual pattern recognition responses of a trained animal. In conversations at this conference, similar observations of 40 Hz oscillation in motor cortex and in the EMG were reported. It thus appears that cortical computation in general may occur by dynamical interaction of resonant modes, as has been thought to be the case in the olfactory system. Given the sensitivity of neurons to the location and arrival times of dendritic input, as emphasized in the work of Rall, the sucessive volleys of pulses that are generated by the collective oscillation of a neural net may be ideal for the formation and reliable longe range transmission of the collective activity of one cortical area to another. The oscillation can serve a macroscopic clocking function and entrain or "bind" the relevant microscopic activity of disparate cortical regions into a well defined phase coherent collective state or "gestalt". This can overide irrelevant microscopic activity and produce coordinated motor output. If this view is correct, then oscillatory network modules form the actual cortical substrate of the diverse sensory, motor, and cognitive operations now studied in static networks. It must ultimately be shown how those functions can be accomplished with oscillatory dynamics.

In particular, we are interested here in modeling category learning and object recognition, *after* feature preprocessing. Equivalence classes of ratios of feature outputs in feature space must be established as prototype "objects" or categories that are invariant over particular sensory instances. Without categories the world never repeats. This is the kind of function generally hypothesized for prepyriform cortex in the olfactory system[9], or inferotemporal cortex in the visual system. In our network module, all inputs in the basin of attraction of a particular periodic attractor are "classified" as being in the same category because they give the same output to the rest of the brain. By the arguments above, the phase locked output of the feature binding process hypothesized for primary visual cortex[8] would be ideal as an input to this object categorization system.

2 Minimal Network

Analog state variables, recurrence, oscillation, and bifurcation into and out of categorizing attractors (during respiration in olfaction) are hypothesized to be essential features of cortical networks which we explore in this approach. Explicit modeling of known excitatory and inhibitory neurons, and use of only known long range connections is also a basic requirement to have a biologically feasible network architecture. We have determined a "minimal" model that is intended to assume the least anatomically justified coupling sufficient to allow function as an oscillatory associative memory. The network is meant only as a cartoon of the real biology, which is designed to reveal the general mathematical principles and mechanisms by which the actual system might function. Such principles can then be observed or applied in other contexts as well.

We take our network units to represent local populations of neurons. Analog state variables then arise naturally as continuous local pulse densities and cell voltage averages. Smooth sigmoidal population input-output functions, whose slope increases with arousal, have been measured in the olfactory system [5]. Long range excitatory to excitatory connections are well known as "associational" connections in olfactory cortex[9] and cortico-cortico

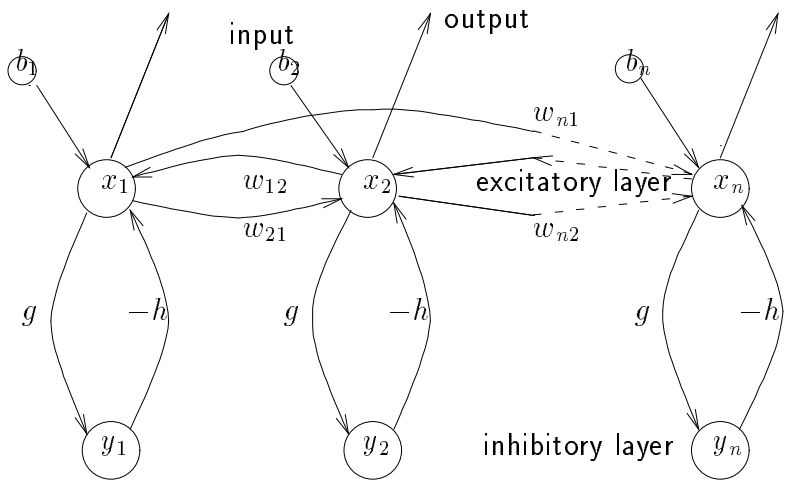

Figure 1: Minimal network of excitatory cell populations x_i, inhibitory cell populations y_i, inputs b_i, adaptive excitatory to excitatory connections W_{ij}, and constant local inhibitory feedback connections g and $-h$.

connections in neocortex. Since our units are neural populations, we can expect that some density of full cross-coupling exists in the system[9], and our weights are taken to be the average synaptic strengths of these connections. At the population level, coupling symmetry in these average connection strenghts can emerge from the operation of a Hebbian learning rule on initially random connections. Local inhibitory "interneurons" are a ubiquitous feature of the anatomy of cortex [8]. It is unlikely that they make long range connections (> 1 mm) by themselves. These connections, and even the debated interconnections between them, are therefore left out of a minimal model. The resulting network of figure 1 is a fair cartoon of the well studied circuitry of prepyriform cortex.

For an N dimensional system, this minimal coupling structure is described mathematically by the matrix T, shown below. W is the $N/2 \times N/2$ matrix of excitatory interconnections, and gI and hI are $N/2 \times N/2$ identity matrices multiplied by the positive scalars g, and h. These give the strength of coupling around local inhibitory feedback loops. A state vector is composed of local average cell voltages for $N/2$ excitatory neuron populations $\vec{x}$ and $N/2$ inhibitory neuron populations $\vec{y}$. Network equations with this coupling are shown below, in component form, along with the matrix T.

$$T = \begin{bmatrix} W & -hI \\ gI & 0 \end{bmatrix} \quad , \qquad \dot{x}_i = -\tau x_i - h y_i + \sum_{j=1}^{N/2} W_{ij}\sigma(x_j) + b_i \tag{1}$$

$$\dot{y}_i = -\tau y_i + g x_i \tag{2}$$

where $\sigma(x) = tanh(x)$ or some other sigmoidal function symmetric about 0. Here we are assuming at the outset that the "squashing" sigmoids of the excitatory interconnections confine the activity of the interactions between inhibitory and excitatory cell populations to a linear region of their axonal nonlinearities. Intuitively, since the inhibitory units receive no direct input and give no direct output, they act as hidden units that create oscillation for

the amplitude patterns stored in the excitatory cross-connections W. This may be viewed as a simple generalization of the analog "Hopfield" network architecture to store periodic instead of static attractors.

If we expand this network to third order in a Taylors series about the origin, we get a network that looks something like,

$$\dot{x}_i = -\tau x_i - h y_i + \sum_{j=1}^{N/2} W_{ij} x_j - \sum_{jkl=1}^{N/2} W_{ijkl} x_j x_k x_l + b_i, \quad (3)$$

$$\dot{y}_i = -\tau y_i + g x_i, \quad (4)$$

where $\sigma'(0) = 1$, and $\frac{1}{3!}\sigma'''(0)(< 0)$ is absorbed into W_{ijkl}. A sigmoid symmetric about zero has odd symmetry, and the even order terms of the expansion vanish, leaving the cubic terms as the only nonlinearity. The actual expansion of the excitatory sigmoids in (1) (in this coordinate system) will only give cubic terms of the form $\sum_{j=1}^{N/2} W_{ij} x_j^3$. The competitive (negative) cubic terms of (3) therefore constitute a more general and directly programmable nonlinearity that is independent of the linear terms.

Normal form theory shows that these cubics are the essential nonlinear terms required to store oscillations, because of the (odd) phase shift symmetry required in the vector field. They serve to create multiple periodic attractors by causing the oscillatory modes of the linear term to compete, much as the sigmoidal nonlinearity does for static modes in a Hopfield network. Intuitively, these terms may be thought of as sculpting the maxima of a "saturation" landscape, into which the modes with positive eigenvalues expand, and positioning them to lie in the directions specified by the eigenvectors to make them stable. We use this network directly as our biological model. From a physiological point of view, (3) may be considered a model of a biological network which is operating in the linear region of the known *axonal* sigmoid nonlinearities, and contains instead these higher order *synaptic* nonlinearities.

3 Learning

A Hebb-like learning rule for the weights of this net has been derived from a more general "projection algorithm" for recurrent analog networks. The projection algorithm [1, 3] specifies learning rules which analytically guarantee content addressable memory storage of different kinds of analog attractors - static, periodic, and chaotic - in the same network. An N node network has N units of capacity. It costs one unit per static attractor, 2 units per Fourier component of each continuous periodic sequence to be stored, and 4 units per chaotic attractor. There are no undesired or "spurious" attractors, and a strict Liapunov function in a special coordinate system governs the approach of initial conditions to the nearest stored trajectory. Latest theoretical work shows, for the network above, that only the excitatory cross connections are required for pattern learning, and that only $(N/2)^2$ of these $(N/2)^4$ possible higher order weights are required in principle to approximate the performance of the projection algorithm. We have shown also that chaotic attractors of dimension < 4, resulting from the interaction of two oscillatory modes, may be stored in this minimal net-

work. Measurements of Freeman have suggested that chaotic attractors of this dimension may occur in the olfactory system[2].

For orthonormal static patterns $\vec{x}^s$, the projection operation for the W matrix reduces to an outer product, or "Hebb" rule, and the projection for the higher order weights becomes a *multiple* outer product rule:

$$W_{ij} = \sum_{s}^{N/2} \alpha^s x_i^s x_j^s \quad , \qquad W_{ijkl} = c\delta_{ij}\delta_{kl} - d \sum_{s=1}^{N/2} x_i^s x_j^s x_k^s x_l^s \ . \tag{5}$$

The first rule is guaranteed to establish desired patterns $\vec{x}^s$ as eigenvectors of the W matrix with corresponding eigenvalues α^s. In the minimal net, these real eigenvalues and eigenvectors (modes) learned for W are converted by the network structure into complex congugate eigenvalues and eigenvectors (modes) for the complete coupling matrix T which includes the local inhibitory feedback. These complex modes are standing wave oscillations whose amplitudes are the magnitudes of the eigenvectors. The frequency w^s of each stored oscillation s is a specified function of the eigenvalue α^s corresponding to its amplitude pattern $\vec{x}^s$ and the strength of the fixed feedback connections g and h within the local oscillators, $\lambda^s_{1,2} = 1/2(\alpha_s \pm \sqrt{\alpha_s^2 - 4hg})$, and $\omega^s = \sqrt{4hg - \alpha_s^2}$.

The second rule, with $c > d$, gives higher order weights W_{ijkl} for the cubic terms in (3) which ensure, by the projection theorem, that the these complex modes of the coupling matrix T become the only attractors in the vector field of the full nonlinear system. The form of the eigenvectors of T and the corresponding asymptotic solutions $X^s(t)$ established by this learning rule are shown below,

$$\begin{bmatrix} |\vec{x}^s| e^{i\vec{\theta}^s_x} \\ \sqrt{\frac{g}{h}} |\vec{x}^s| e^{i\vec{\theta}^s_y} \end{bmatrix} \quad \Rightarrow \quad X^s(t) = \begin{bmatrix} |\vec{x}^s| e^{i\vec{\theta}^s_x + i\omega^s t} \\ \sqrt{\frac{g}{h}} |\vec{x}^s| e^{i\vec{\theta}^s_y + i\omega^s t} \end{bmatrix} .$$

Because of the restricted coupling, the only oscillations possible in this network have zero phase lag across space, i.e. are standing waves. The phase θ_x, θ_y is constant over the components of each kind of neural population x and y, and differs only between them. This is basically what is observed in the olfactory bulb (primary olfactory cortex) and prepyriform cortex[2]. The phase of inhibitory components θ_y in the bulb lags the phase of the excitatory components θ_x by approximately 90 degrees. It is easy to have α's and feedback weights g and h in this model to get phase lags of nearly 90 degrees. This analysis holds as well with minor variations for a network with adaptive excitatory to inhibitory cross-coupling (W in the lower left of T), which may constitute a plausible model of the olfactory bulb as an associative memory.

4 Higher Order Synapses

Adding the higher order weights corresponds, in connectionist language, to increasing the complexity of our neural population nodes to become "higher order" or "sigma-pi" units[10]. Figure 2 shows schematically how subnetworks within a population unit can compute products of the activities on incomming primary connections W_{ij}, during higher order Hebbian

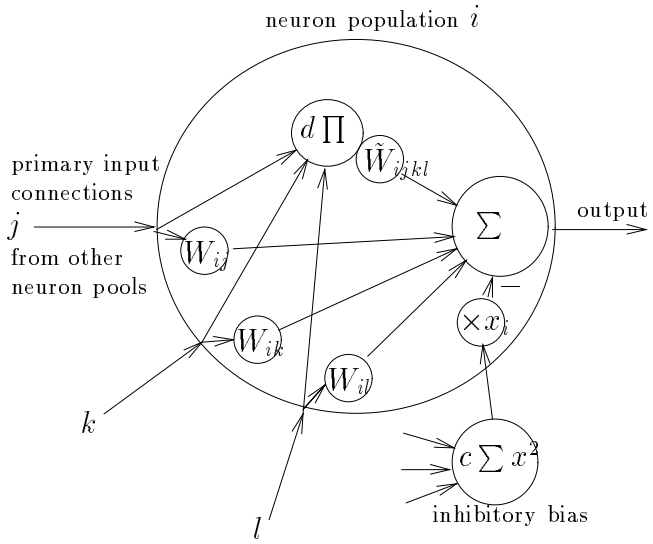

Figure 2: Neural population subnetwork acting as a sigma-pi unit. It uses secondary higher order synaptic weights $\tilde{W}_{ijkl}$ on products of the activities of incoming primary connections W_{ij}, and receives a global inhibitory bias.

learning, to establish a weight W_{ijkl}. These secondary higher order synapses are then used in addition to the synapses W_{ij}, during operation of the overall network, to weight the effect of triple products of inputs in the output summation of the subnet.

When the Hebbian learning rule (5) is used, the higher order weights W_{ijkl} of the network model (3) can be decomposed so that (3) becomes,

$$\dot{x}_i \;=\; -\tau x_i - h y_i + \sum_{j=1}^{N/2} W_{ij} x_j + d \sum_{jkl=1}^{N/2} \tilde{W}_{ijkl} x_j x_k x_l - c x_i \sum_{j=1}^{N/2} x_j^2 + b_i \qquad (6)$$

where $\tilde{W}_{ijkl}$ comes from the multiple outer product in (5), and $-cx_i \sum_{j=1}^{N/2} x_j^2$ comes from the $c\delta_{ij}\delta_{jk}$ term in (5). This single weight negative term corresponds to a shunting inhibitory bias which depends on the global excitatory activity of the network. "Shunting" here means multiplied by the current average cell voltage x_i of population i. This is an input which is identical for all excitatory neural populations and could be calculated by a single node of the network which receives input from all excitatory populations. Such a node might correspond to one of the nuclei which lie below the prepyriform cortex. These send and receive the diffuse projections required to and from prepyriform cortex. The constants c and d of eq. (6) give the magnitude of the inhibitory bias c and the average higher order weight d. Since $c > d$ guarantees stability of all stored patterns, the greater the bias c relative to d, the more robust is the stability of the present attractor, and vice versa. This is a possible mechanism for central control of attractor transitions (bifurcations) related perhaps to selective attention.

Using the long range excitatory connections W_{ij} available, some number of the higher order synaptic weights W_{ijkl} could be realized within a neural population in the axo-dendritic

interconnection plexus known as "neuropil". This a dense feltwork of tiny fibers. Single axon fibers are known to bifurcate into multiple branches that contribute many separate synapses to the dendrites of target cells. It is also well known that neighboring synapses on a dendrite can interact in a nonlinear fashion that has been modeled as higher order synaptic terms by some researchers. It has been suggested that the neuropil may be dense enough to allow the crossing of every possible combination of jkl axons in the vicinity of some dendritic branch of at least one neuron in neuron pool i. Trophic factors stimulated by the coactivation of the axons and the dendrite could cause these axons to form a "cluster" of nearby synapses on the dendrite to realize a jkl product synapse[10], or clusters formed randomly during development could be selected for survival by coactivation during later learning experience.

References

[1] B. Baird. A bifurcation theory approach to vector field programming for periodic attractors. In *Proc. Int. Joint Conf. on Neural Networks, Wash. D.C.*, pages I381–388, June 18 1989.

[2] B. Baird. Bifurcation and learning in network models of oscillating cortex. *Physica D*, 42:365–384, 1990.

[3] B. Baird. A learning rule for cam storage of continuous periodic sequences. In *Proc. Int. Joint Conf. on Neural Networks, San Diego*, pages III493–498, June 1990.

[4] B. Baird. *Bifurcation Theory Approach to the Analysis and Synthesis of Neural Networks for Engineering and Biological Modeling.* Research Notes in Neural Computing. Springer, 1991. to appear.

[5] H. Eeckman, F. The sigmoid nonlinearity in prepyriform cortex. In D. Anderson, editor, *Neural Information Processing Systems*, page 135, New York, 1988. American Institute of Physics.

[6] W. Freeman. *Mass Action in the Nervous System.* Academic Press, New York, 1975.

[7] W. Freeman and B. Baird. Relation of olfactory eeg to behavior: Spatial analysis. *Behavioral Neuroscience*, 101:393, 1987.

[8] C. M. Gray and W. Singer. Stimulus dependent neuronal oscillations in the cat visual cortex area 17. *Neuroscience [Suppl]*, 22:1301P, 1987.

[9] L. B. Haberly and J. M. Bower. Olfactory cortex: model circuit for study of associative memory? *Trends in Neuroscience*, 12(7):258, 1989.

[10] B. Mel and C. Koch. Sigma-pi learning: On radial basis functions and cortical associative learning. In D. Touretzky, editor, *Advances in Neural Information Processing 2*, page 474. Morgan Kaufman, 1990.